II Encontro Galaico-Português de Biometria
Santiago de Compostela, 30 de Junho, 1 e 2 de Julho de 2016

A Periodic State Space Model to Monthly Long-term Temperature Data

Marco Costa[1,2] and Magda Monteiro[1,2]

[1]ESTGA - School of Technology and Management of Águeda, University of Aveiro, Portugal
[2]CIDMA - Center for Research & Development in Mathematics and Applications, University of Aveiro, Portugal

ABSTRACT

This work presents a periodic state space model to model monthly temperature data. Additionally, some issues are discussed, as the parameter estimation or the Kalman filter recursions adapted to a periodic model. This framework is applied to monthly long-term temperature time series of Lisbon.

Key-words: State space model, Kalman filter, periodic data, temperature.

1. INTRODUCTION

The rise in global temperature has been an increasing concern of several authorities. According to the Intergovernmental Panel on Climate Change, the world's greenhouse gas emissions are still increasing and on the present path, the global temperature rise will substantial exceed the limit goal of two degrees Celsius that countries have agreed upon in order to avoid the most dangerous impacts of climate change. In the United Nations (UN) Framework on Climate Change, it was adopted the Paris Agreement, on 12 December 2015, at the UN Climate Change Conference in France where parties committed to take ambitious actions to keep global temperature rise below 2 degrees Celsius by the end of the century [9].
Climate change monitoring is performed with multidisciplinar approaches. However, the statistical analysis has been a key in this theme. Usually, monitoring climate change is performed based on regional or global temperature long-term data. Indeed, temperature data is considered in several studies of climate change or as the main variable under analysis or as a covariate on modeling other environmental variables, for instance in rainfall modeling. Thus, long-term temperature time series are, par excellence, the most appropriate data to study climate change, mainly in the global or local warming.
Indeed, modeling and foresight scenarios of environmental and climate variables are based on the selection of climate models using the analysis of other variables as rainfall and temperature [5]. On the other hand, the impact of changes in environmental variables changes, as the temperature or the precipitation, on environmental issues have been assessed and analyzed in order to explore the implications under a future climate [3].
The most applied statistical models in order to analyze climate time series, in particular of temperatures data sets, are the linear models and the usual time series models as ARIMA models [1, 2, 4]. These models are largely known as simple and they have well-known optimal properties under the usual assumptions (normality of errors, etc.). Some extensions and combinations of these models

have been proposed in environmental data analysis in order to improve models performance and to increase the predictions accuracy.

State space models have been largely considered an extension of the usual linear models which incorporate a versatile stochastic structure at the same time that integrate temporal dependence. The goal of this work is to analyze monthly long-term temperature time series of the Portuguese city of Lisbon. In [7] propose and implement a methodology for detection and correction of series of monthly temperatures for Lisbon, Oporto and Coimbra, providing the homogenization of these Portuguese long-term temperature data series (data in [8]). For instance, see Figure 1 related to Lisbon homogenized monthly temperature data.

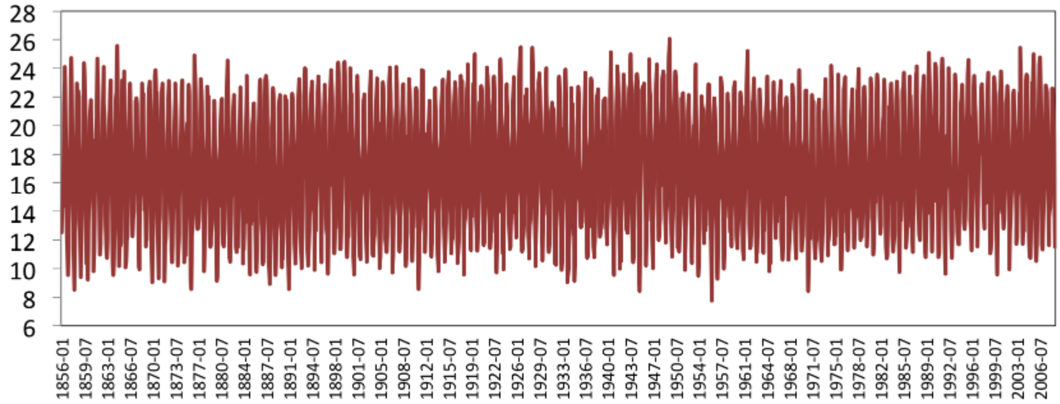

Figure 1: Lisbon homogenized monthly temperature data.

State space models have been largely applied in several areas of applied statistics. In particular, the linear state space models have desirable properties and they have a huge potential in time series modeling that incorporates latent processes.

Once a model is placed in the linear state space form, the most usual algorithm to predict the latent process, the state, is the Kalman filter algorithm. This algorithm is a procedure for computing, at each time t $(t = 1, 2, \ldots)$, the optimal estimator of the state vector based on the available information until t and its success lies on the fact that is an online estimation procedure.

The main goal of the Kalman filter algorithm is to find predictions for the unobservable variables based on observable variables related to each other through a set of equations forming the state space model. Indeed, in the context of linear state space models, the Kalman filter produces the best linear unbiased estimators. When the errors and the initial state are Gaussian, the Kalman filter predictors are the best unbiased estimators in the sense of the minimum mean square error. However, the optimal properties only can be guaranteed when all model's parameters are known [6].

2. THE PERIODIC STATE SPACE MODEL - PSSM

In order to accommodate data properties as temporal correlation and the periodic behavior, a periodic state space model is proposed. This model is defined as

$$
\begin{aligned}
Y_{s,n} &= [S(n-1)+s]X_{s,n} + D_{s,n} \quad + e_{s,n} &(1)\\
X_{s,n} &= \mu_s + \quad _s(X_{s-1,n} - \mu_{s-1}) + \quad _{s,n} &(2)
\end{aligned}
$$

where

- s is the season of the year with $s = 1, 2, ..., S$;
- n is the year with $n = 1, 2, ..., N$;
- $Y_{s,n}$ represents the time series observation in the s^{th} season of the n^{th} year, that is, the $[S(n-1)+s]^{\text{th}}$ observation of the time series;

- $= [\ _1\ _2 \ldots\ _S]$ are unknown parameters representing the fixed effects in the model;
- $D_{s,n}$ is an $1 \times S$ matrix of known values, a design matrix;
- the error process $(e_{s,n})$ is a white noise disturbance, which is assumed to have $var(e_{s,n}) = \ ^2_e$;
- μ_s is the mean of the process $(X_{s,n})$ for the s^{th} season, $\ _s$ is the autoregressive parameter for season s and $\ _{s,n}$ is the white noise disturbance;
- the error process $(\ _{s,n})$ are distributed IID$(0, \ ^2_s)$ and it is assumed to have an covariance defined by

$$cov(\ _{s,n},\ _{s-i,n}) = \begin{matrix} \ ^2_{,s} & i = 0 \\ 0 & i = 0 \text{ for } i = 1, 2, ..., S. \end{matrix}$$

The first equation is referred as the observation equation and the second is called the state equation.

4. GAUSSIAN LIKELIHOOD ESTIMATION

Under the assumptions that the initial state, the state noise $\ _{s,n}$ and the observation noise $e_{s,n}$ are mutually independent and normally distributed, for a given realization $\mathbf{Y} = (Y_{1,1}, Y_{2,1}, ..., Y_{S,N})$ the logarithm of the Gaussian likelihood function is computed as follows

$$\log \mathcal{L}(\ ; \mathbf{Y}) = -\frac{NS}{2} \log(2\ ) - \frac{1}{2} \sum_{n=1}^{N} \sum_{s=1}^{S} \log(\ _{s,n}) - \frac{1}{2} \sum_{n=1}^{N} \sum_{s=1}^{S} \frac{\ ^2_{s,n}}{\ _{s,n}}$$

where $\ = (\ ,\ ^2_e, \mu_1, ..., \mu_S,\ _1, ...,\ _S,\ ^2_{1,}\ , ...,\ ^2_{S,}\ )$, $\ _{s,n} = Y_{s,n} - Y_{s-1,n}$ is the innovation and $\ _{s,n}$ is the mean square error of the innovation. The forecast one-step-ahead of $Y_{s,n}$, $Y_{s-1,n}$, is obtained by the Kalman filter equations reformulated to the model (1)-(2). The maximum likelihood estimates are obtained through numerical procedures.

5. SOME RESULTS

The modeling procedure and the Kalman filter algorithm allow obtaining one-step-ahead forecasts and some overall results. In the first case, Figure 2 shows the one-step-ahead forecasts in the last decade of the period analyzed in Lisbon. It seems that the model has a good adjustment to the real data.

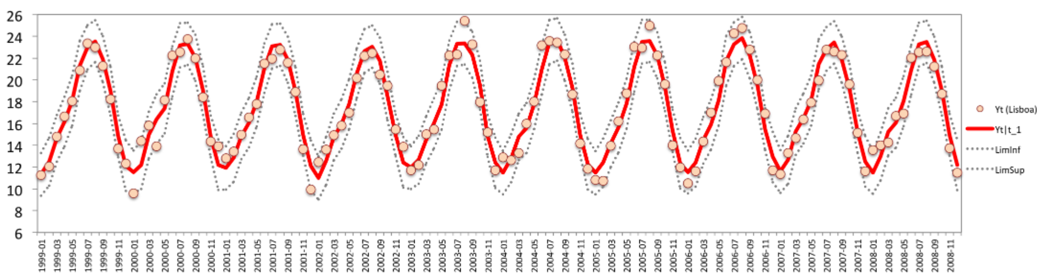

Figure 2: One-step-ahead forecasts in the last decade of the period analyzed in Lisbon.

On the other hand, the model's parameters allow to concluded that there is a global temperature rise. However, the months with the highest average rates of temperature increase by century are January (+0,9ºC), March (+1,1ºC), October (+0,8ºC) and December (+0,8ºC).

ACKNOWLEDGEMENTS

Authors were partially supported by Portuguese funds through the CIDMA - Center for Research and Development in Mathematics and Applications, and the Portuguese Foundation for Science and Technology ("FCT– Fundação para a Ciência e a Tecnologia"), within project UID/MAT/04106/2013.

References

[1] Alpuim, T., El-Shaarawi, A. (2009). Modeling monthly temperature data in Lisbon and Prague. *Environmetrics* 20 835–852.

[2] Bližňák, V., Valente, M.A., Bethke, J.E. (2015). Homogenization of time series from Portugal and its former colonies for the period from the late 19th to the early 21st century. *International Journal of Climatology* 35 2400–2418.

[3] Flannigan, M.D., Wotton, B. M., Marshall, G.A., Groot, W.J., Johnston, J., Jurko, N., Cantin, A.S. (2016). Fuel moisture sensitivity to temperature and precipitation: climate change implications *Climatic Change* 134 59–71.

[4] Freitas, C.R., Dedekind, M.O., Brill, B.E. (2015). A Reanalysis of Long-Term Surface Air Temperature Trends in New Zealand *Environmental Modeling & Assessment* 20 399-410.

[5] Guerreiro, S.B., Kilsby, C. G., Fowler, H.J. (2016). Rainfall in Iberian transnational basins: a drier future for the Douro, Tagus and Guadiana? *Climatic Change* DOI:10.1007/s10584-015-1575-z.

[6] Harvey, A.C. (1996). *Forecasting Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.

[7] A.L. Morozova, M.A. Valente, Homogenization of Portuguese long-term temperature data series: Lisbon, Coimbra and Porto, *Earth Syst. Sci. Data* **4**: 187–213 (2012).

[8] A.L. Morozova, M.A. Valente, Homogenization of Portuguese long-term temperature data series: Lisbon, Coimbra and Porto, doi:10.1594/PANGAEA.785377 (2012).

[9] United Nations (2016). http://www.un.org/sustainabledevelopment/climate-change/ (accessed in January 22, 2016).